

Perbandingan Metode K-Nearest Neighbor Dan Naïve Bayes Untuk Klasifikasi Gender Berdasarkan Mata

Calvin Kurniawan¹, Hafiz Irsyad²

^{1,2}Universitas Multi Data Palembang; Jalan Rajawali No. 14 Palembang, 0711-376400

³Prodi Informatika, FIKR, Universitas Multi Data Palembang

e-mail: *¹calvinsants@mhs.mdp.ac.id, ²HafizIrsyad@mdp.ac.id

Abstrak

Mata merupakan organ penglihatan, yang terletak di rongga orbital. Bentuknya bulat, sekitar 2,5cm. Ruang antara mata dan orbit diisi oleh jaringan lemak. Dinding tulang dan lemak orbit yang dapat melindungi mata dari terluka. Secara struktural, kedua mata terpisah, tetapi mereka adalah pasangan. Hal Dapat dilihat hanya dengan satu mata, tetapi penglihatan tiga dimensi Merasa tidak nyaman Ketika hanya menggunakan satu mata, terutama dalam hubungan penghakiman jarak. Gender adalah penggolongan secara gramatikal terhadap kata-kata dan kata-kata lain yang berkaitan dengannya yang secara garis besar berhubungan dengan keberadaan dua jenis kelamin atau kenetralan. Penelitian ini bertujuan untuk klasifikasi Gender berdasarkan mata menggunakan metode K-NN dan Naïve Bayes. Data yang dibagi menjadi dua yaitu mata pria dan mata Wanita adalah 10.541 data uji dan 4518 data latih dilanjutkan dengan grayscale, HOG, HSV, Cropping. Pada metode k-NN dengan fitur HSV dan HOG dengan Cropping metode K-NN menghasilkan akurasi yang lebih tinggi di bandingkan Naïve Bayes yaitu sebesar 68,61%.

Kata kunci — Mata, KNN, Naïve Bayes, Gender, Cropping

Abstract

The eye is an organ of vision, which is located in the orbital cavity. The shape is round, about 2.5 cm. The space between the eye and the orbit is filled with fat tissue. The bony and fatty walls of the orbit protect the eye from injury. Structurally, the eyes are separate, but they are a pair. Things Can be seen with only one eye, but three-dimensional vision Feel uncomfortable when using only one eye, especially in relation to distance judgment. Gender is a grammatical classification of words and other related words which are generally related to the existence of two sexes or neutrality. This study aims to classify gender based on eyes using the K-NN and Naïve Bayes methods. The data which is divided into two, namely male eyes and female eyes are 10,541 test data and 4518 training data followed by grayscale, HOG, HSV, Cropping. In the k-NN method with HSV and HOG features with Cropping the K-NN method produces a higher accuracy than Naïve Bayes, which is 68.61%.

Keywords— Eye, KNN, Naïve Bayes, Gender, Cropping

1. PENDAHULUAN

Mata adalah organ yang diciptakan oleh Tuhan dan merupakan salah satu organ terpenting. Orang hanya bisa mendapatkan hingga 80% dari informasi menurut (Kurmasela et al. 2013).

Histogram of Oriented Gradient (HOG) adalah sebuah metode yang digunakan dalam Image Processing yang bertujuan untuk mendeteksi objek. Teknik ini digunakan untuk menghitung nilai *gradient* dalam daerah tertentu pada suatu *Image* atau gambar.

Hue Saturation Value (HSV) adalah definisi dari *Hue*, *Saturation*, *Value*. *Hue* adalah ukuran dari jenis warna yang ada, artinya warna menjadi lebih berwarna. *Saturation* artinya nilai saturasi bertambah seiring dengan bertambah gelapnya warna. *Value* artinya nilai saturasi berkurang, dan nilai itu adalah nilai kecerahan. berwarna. warna.

Gender sering diartikan sebagai jenis kelamin. *Gender* adalah klasifikasi gramatikal terhadap kata-kata dan kata-kata terkait lainnya yang berkaitan dengannya yang secara garis besar berhubungan dengan keberadaan dua jenis kelamin atau kenetralan. [1] *Gender* juga bersangkutan dengan beda perilaku peran, fungsi dan tanggung jawab laki-laki dan perempuan sebagai hasil kesepakatan atau hasil bentukan masyarakat. *Gender* juga merupakan suatu konstruksi budaya yang sifatnya terbuka terhadap bagi segala perubahan [2].

Metode yang digunakan untuk membandingkan antara metode *K-NN* dan *Naïve Bayes* dengan menggunakan mata untuk membedakan gender. *K-NN* merupakan metode klasifikasi yang menentukan kategori berdasarkan mayoritas kategori pada *k-Nearest Neighbors* [3] sedangkan *Naïve Bayes Classifier* (NBC) adalah metode pada probabilistic reasoning yang menghitung sekumpulan probabilitas dengan menambahkan kombinasi frekuensi dan nilai dalam kumpulan data yang diberikan. Tujuan dari penelitian ini adalah untuk mengetahui perbandingan tingkat akurasi klasifikasi Gender dengan menggunakan *KNN* (*K-Nearest Neighbors*) dan *Naïve Bayes* pada citra mata serta menerapkan metode *KNN* (*K-Nearest Neighbors*) dan *Naïve Bayes* terhadap citra mata .

2. METODE PENELITIAN

2.1 *KNN (K-Nearest Neighbors)*

K-NN merupakan metode klasifikasi yang menentukan kategori berdasarkan mayoritas kategori pada *k-Nearest Neighbors* [3]. Klasifikasi didasarkan pada objek latih yang paling dekat dengan objek uji. *k-NN* dapat menangani data campuran dalam bentuk polinomial dan numerik [4]. Tahapan proses klasifikasi *k-NN* adalah sebagai berikut [3]:

1. Tentukan nilai *k*.
2. Hitung jarak antara data uji dan semua data latih.
3. Perhitungan jarak antara data uji dengan semua data latih diurutkan berdasarkan jarak terkecil.
4. Bentuklah kelompok berdasarkan nilai dengan nilai jarak.
5. Pilih nilai (kategori) yang paling sering muncul.

Cara kerja metode *k-NN* adalah menentukan kelas objek baru tersebut dengan mengukur jarak antara objek baru dengan objek lama [5]. Jaringan Saraf Relatif tidak sensitif terhadap kesalahan dalam kumpulan data dan dapat digunakan untuk memproses kumpulan data yang besar [6], [5] [4]

2.2 *Naïve Bayes Classifier*

Naïve Bayes Classifier (NBC) adalah metode pada probabilistic reasoning yang menghitung sekumpulan probabilitas dengan menambahkan kombinasi frekuensi dan nilai dalam kumpulan data yang diberikan. NBC adalah algoritma klasifikasi yang sangat efektif (untuk mendapatkan hasil yang benar) dan efisien (proses penalaran menggunakan input yang ada dengan cara yang relatif cepat). Algoritma NBC dirancang untuk mengklasifikasikan kategori data tertentu.

Kinerja classifier diukur dengan predictive accuracy [7]. Keuntungan lain dari NBC adalah dapat menangani data diskrit dan kontinu. Pada proses pencarian kategori terbaik pada saat data diskrit, jika diberikan k atribut independen maka nilai probabilitasnya dapat seperti pada Persamaan 1.

$$P(x_1, \dots, x_k | C) = P(x_1 | C) \times \dots \times P(x_k | C) \quad (1)$$

Keterangan

P = Probabilitas

c = kelas C

x = atribut I

awal k = atribut I akhir

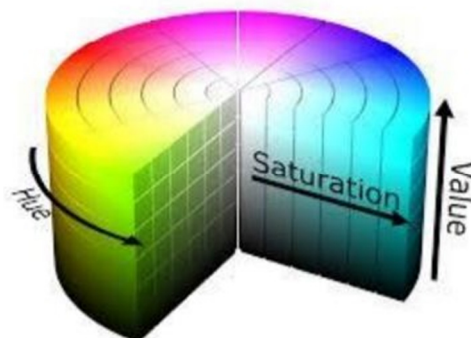
Jika atribut ke-i bersifat diskrit atau kategori, maka $P(x_i | C)$ di estimasi sebagai frekuensi relatif sampel yang memiliki nilai x_i sebagai atribut ke-i dalam kelas C.

2.3 Histogram Of Gradients (HOG)

Histogram Of Gradients (HOG) adalah sebuah metode yang di gunakan dalam image processing untuk bertujuan deteksi objek. Teknik ini menghitung nilai gradient dalam daerah tersebut pada sesuatu image. Tiap image mempunyai karakteristik yang di tunjukan oleh distribusi gradient. Karakteristik ini diperoleh dengan membagi image kedalam daerah kecil yang disebut cell. Tiap cell disusun sebuah histogram dari sebuah gradient t. Kombinasi dari histogram ini dijadikan sebagai descriptor yang mewakili sebuah obyek.

2.4 Hue Saturation Value (HSV)

HSV adalah definisi dari hue, saturation, value. Hue adalah ukuran jenis warna yang ada, Saturation berarti nilai saturasi bertambah dengan bertambahnya jumlah warna, tetapi fading berarti nilai saturasi berkurang, dan value adalah nilai lightness suatu warna [8]. Dalam HSV, warna adalah sudut antara 0 dan 360 yang mewakili jenis warna dalam spektrum warna, saturasi mewakili kemurnian atau tingkat abu-abu suatu warna dan value adalah ukuran kecerahannya. Kesimpulannya adalah bahwa warna HSV ini model meliputi Hue, Saturation, dan Value. Ada tiga komponen penting dari HSV, dan ketiganya berperan, Hue adalah warna yang mengukur panjang gelombang warna, Saturation mengacu pada kemurnian warna, Value mengacu pada tingkat kecerahan warna [9].



Gambar 1. Hue Saturation Value

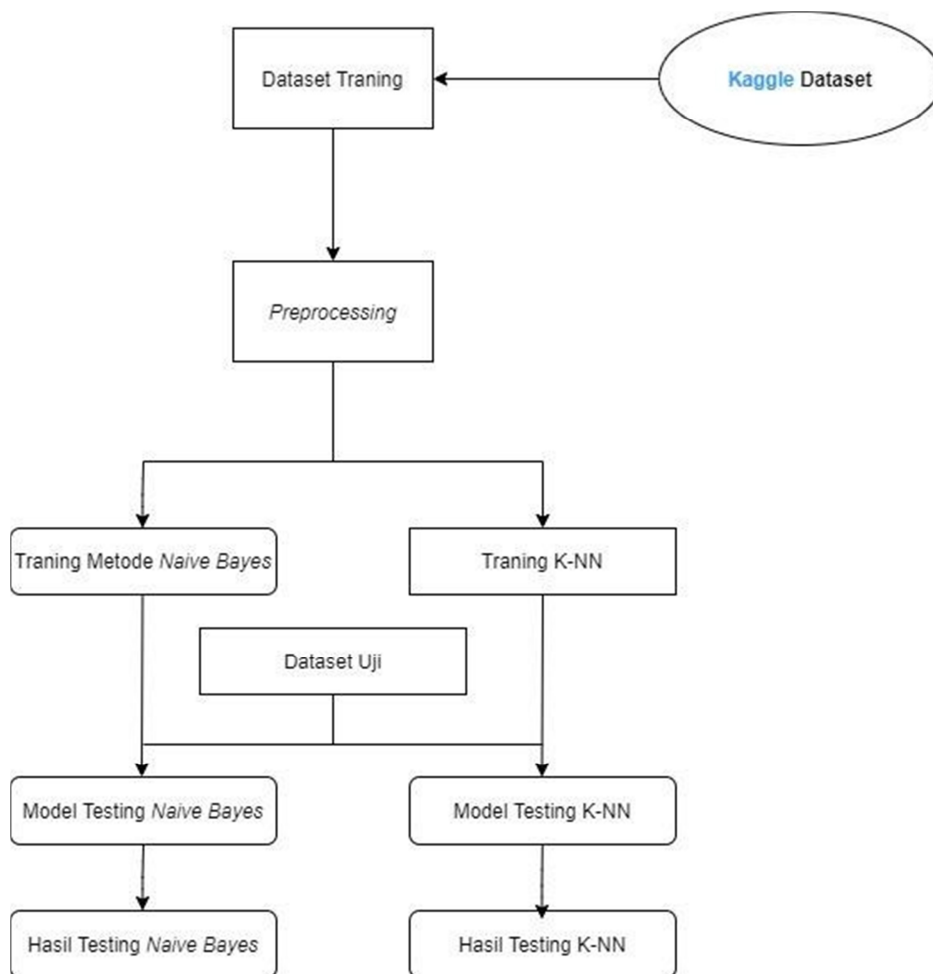
Model warna HSV didapat dari hasil turunan model warna RGB, oleh karena itu untuk mendapatkan model warna HSV dapat dilakukan dengan mengkonversi warna dari RGB ke HSV.

2.5 Perancangan

Pada tahapan ini dilakukan perancangan sistem yang dibutuhkan, Diawali dengan input dataset citra kemudian dilakukan proses klasifikasi Gender dalam skenario pengujian menggunakan metode KNN dan Naïve Bayes. Penelitian ini dilakukan dalam Skema perancangan dapat dilihat pada Gambar 5.

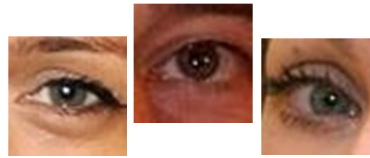
2.5.1 Pengumpulan Data Set

Pada tahap ini, melakukan pengumpulan data berupa dataset Female and Male Eyes berasal dari <https://www.kaggle.com/pavelbiz/eyes-rtte> merupakan dataset bersifat publik.



Gambar 5. Tahap Perancangan

Dataset yang berjumlah 15.060 terdiri dari 8090 jenis mata laki-laki dan 6969 jenis mata perempuan, meliputi 10.541 (70 Persen terdiri dari 5663 mata laki-laki dan 4879) data uji, 4518 data latih (30 Persen terdiri dari 2427 mata laki-laki dan 2091).



Gambar 6. Dataset Mata

2.5.2 Implementasi dan Training Data

Pada tahapan ini dilakukan implementasi dari penelitian dan sistem yang telah dirancang sebelumnya dengan data training yang telah ada sebelumnya agar sistem dapat mengenali dan dapat melakukan klasifikasi terhadap data uji yang telah dikumpulkan sebelumnya

2.8.3 Pengujian

Pada tahap ini sistem akan melakukan pengujian terhadap data uji yang telah diperoleh dan diimplementasikan pada sistem yang telah dibuat, hasil dari klasifikasi tersebut akan dicatat untuk dilakukan analisa untuk mendapatkan *accuracy*, *precision*, *recall* pada tahap selanjutnya.

2.5.3 Evaluasi Hasil Pengujian

Setelah tahapan uji coba, dilakukan analisa hasil pengujian sebelumnya. Hasil uji coba tersebut dihitung untuk mendapatkan tingkat keberhasilan metode yang digunakan dengan *Confusion Matrix*. *Confusion Matrix* adalah salah satu metode yang digunakan untuk mengevaluasi metode-metode klasifikasi. Tabel 3.1 merupakan gambaran sederhana tentang istilah *confusion matrix* dalam keluaran klasifikasi [10].

Tabel 1. Istilah Confusion Matrix

		Kelas Prediksi (Observation)	
		Positif	Negatif
Kelas Aktual (Expeptation)	Positif	TP	FN
	Negatif	FP	TN

Sumber: Akhmad Robim, 2019

True Positive (TP) adalah data yang diklasifikasi dengan tepat sebagai keluaran positif atau benar. Nilai *True Negative* (TN) adalah data yang diklasifikasi dengan tepat sebagai keluaran negative atau salah. *False Positive* (FP) adalah data yang diklasifikasi dengan kurang tepat apabila hasilnya berupa positif atau benar. *False Negative* (FN) adalah data yang diklasifikasi dengan tepat. Berikut dapat dilihat untuk menghitung nilai *precision*, *recall*, dan *accuracy*

3. HASIL DAN PEMBAHASAN

3.1 Skenario k-NN Dengan Ekstraksi HSV dan HOG k=3 Tanpa Cropping

Pengujian ini melakukan tahap ekstraksi HSV dan HOG dengan k=3 tanpa cropping menggunakan model k-NN. Hasil training yang sudah dilakukan memperoleh confusion matrix dapat dilihat pada Tabel 2.

Tabel 2, Confusion Matrix k-NN k=3 Tanpa Cropping

		Actual	
		Female	Male
Predicted	Female	1223	893
	Male	902	1542

Berdasarkan tabel 2 dapat dilihat bahwa jumlah data positif yang terklasifikasi benar oleh sistem TP (*True Positive*) memiliki hasil sebesar 1223 dan terklasifikasi salah oleh sistem FP (*False Positive*) memiliki hasil sebesar 902. Adapun jumlah data negatif yang terklasifikasi salah oleh sistem TN (*True Negative*) memiliki hasil sebesar 1542 dan yang terklasifikasi benar oleh sistem FN (*False Negative*) memiliki hasil sebesar 893. Hasil perhitungan *accuracy*, *precision*, dan *recall*, dapat dilihat pada Tabel 3.

Tabel 3. Hasil Pengujian k-NN Dengan k=3 dan Tanpa Cropping

No.	Class	Precision	Recall
1	Female	57,55%	63,32%
2	Male	57,79%	63,09%
Rata - rata accuracy			60,63%

3.2 Skenario k-NN Dengan ekstraksi fitur HSV dan HOG k=3 Dengan Cropping

Pengujian ini melakukan tahap ekstraksi HSV dan HOG dengan k=3 dengan cropping menggunakan model k-NN. Hasil training yang sudah dilakukan memperoleh confusion matrix dapat dilihat pada Tabel 4.

Tabel 4. Confusion Matrix k-NN Dengan k=3 dan Dengan Cropping

		Actual	
		Female	Male
Female		1403	713
Male		709	1735

Berdasarkan tabel 4 dapat dilihat bahwa jumlah data positif yang terklasifikasi benar oleh sistem TP (*True Positive*) memiliki hasil sebesar 1403 dan terklasifikasi salah oleh sistem FP (*False Positive*) memiliki hasil sebesar 709. Adapun jumlah data negatif yang terklasifikasi salah oleh sistem TN (*True Negative*) memiliki hasil sebesar 1735 dan yang terklasifikasi benar oleh sistem FN (*False Negative*) memiliki hasil sebesar 713. Hasil perhitungan *accuracy*, *precision*, dan *recall* dapat dilihat pada Tabel 5.

Tabel 5. Hasil Pengujian k-NN Dengan k=3 dan Dengan Cropping

No.	Class	Precision	Recall
1	Female	66,43%	66,30%
2	Male	70,87%	70,99%
Rata - rata accuracy			68,81%

3.3 Skenario Naïve Bayes Dengan Ekstraksi HSV dan HOG Tanpa Cropping

Pengujian ini melakukan tahap ekstraksi HSV dan HOG tanpa Cropping menggunakan model Naïve Bayes. Hasil training yang sudah dilakukan memperoleh confusion matrix dapat dilihat pada Tabel 6.

Tabel 6. Confusion Matrix Naïve Bayes Tanpa Cropping

		Actual	
		Female	Male
Predicted	Female	1318	798
	Male	1188	1256

Dari tabel 6 dapat dilihat bahwa jumlah data positif yang terklasifikasi benar oleh sistem (True Positive) memiliki hasil sebesar 1318 dan terklasifikasi salah oleh sistem (False Positive) memiliki hasil sebesar 1188. Adapun jumlah data negatif yang terklasifikasi salah oleh sistem (True Negative) memiliki hasil sebesar 1256 dan yang terklasifikasi benar oleh sistem (False Negative) memiliki hasil sebesar 798. Hasil perhitungan accuracy, precision, dan recall dapat dilihat pada Tabel 7.

Tabel 7. Hasil Pengujian Naïve Bayes Tanpa Cropping

No.	Class	Precision	Recall
1	Female	52,59%	61,14%
2	Male	62,28%	51,39%
Rata - rata accuracy			56,44%

3.4 Skenario Naïve Bayes Dengan Ekstraksi Fitur HSV dan HOG Dengan Cropping

Pengujian ini melakukan tahap ekstraksi HSV dan HOG dengan Cropping menggunakan model Naïve Bayes. Hasil training yang sudah dilakukan memperoleh confusion matrix dapat dilihat pada Tabel 8.

Tabel 8. Confusion Matrix Naïve Bayes Dengan Cropping.

		Female	Male
Female	1305	811	
Male	895	1549	

Dari tabel 8 dapat dilihat bahwa jumlah data positif yang terklasifikasi benar oleh sistem (*True Positive*) memiliki hasil sebesar 1305 dan terklasifikasi salah oleh sistem (*False Positive*) memiliki hasil sebesar 895. Adapun jumlah data negatif yang terklasifikasi salah oleh sistem (*True Negative*) memiliki hasil sebesar 1549 dan yang terklasifikasi benar oleh sistem (*False Negative*) memiliki hasil sebesar 811. Hasil perhitungan *accuracy*, *precision*, *recall* dapat dilihat pada Tabel 9.

Tabel 9. Hasil Pengujian *Naïve Bayes* Dengan *Cropping*

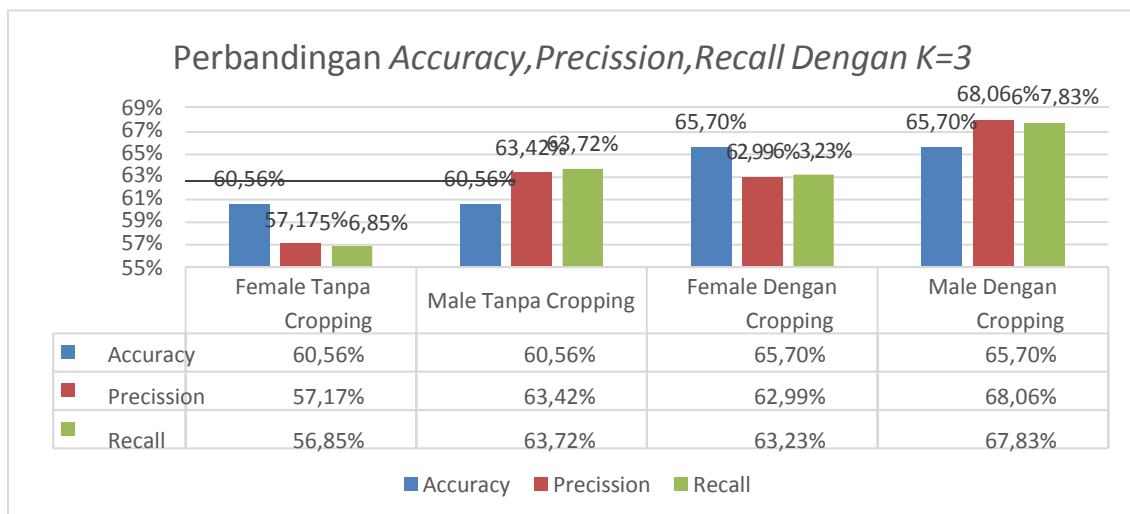
No.	Class	Precision	Recall
1	Female	59,31%	61,67%
2	Male	65,63%	63,37%
Rata - rata accuracy			62,58%

3.2 Analisis Hasil Pengujian

Berdasarkan hasil training pada setiap skenario dapat dilihat bahwa perbandingan hasil klasifikasi menggunakan *Cropping* dan tidak menggunakan *Cropping* serta penggunaan memiliki dampak pengaruh tingkat akurasi pada 2 metode k-NN dan *Naïve Bayes*

3.2.1 Analisis Skenario k-NN Dengan k=3

Pada skenario k-NN yang dilakukan dengan menggunakan k=3. tanpa dan dengan menggunakan *Cropping* untuk klasifikasi mata. Perbandingan *accuracy*, *precision*, dan *recall* dapat dilihat pada Gambar 6.

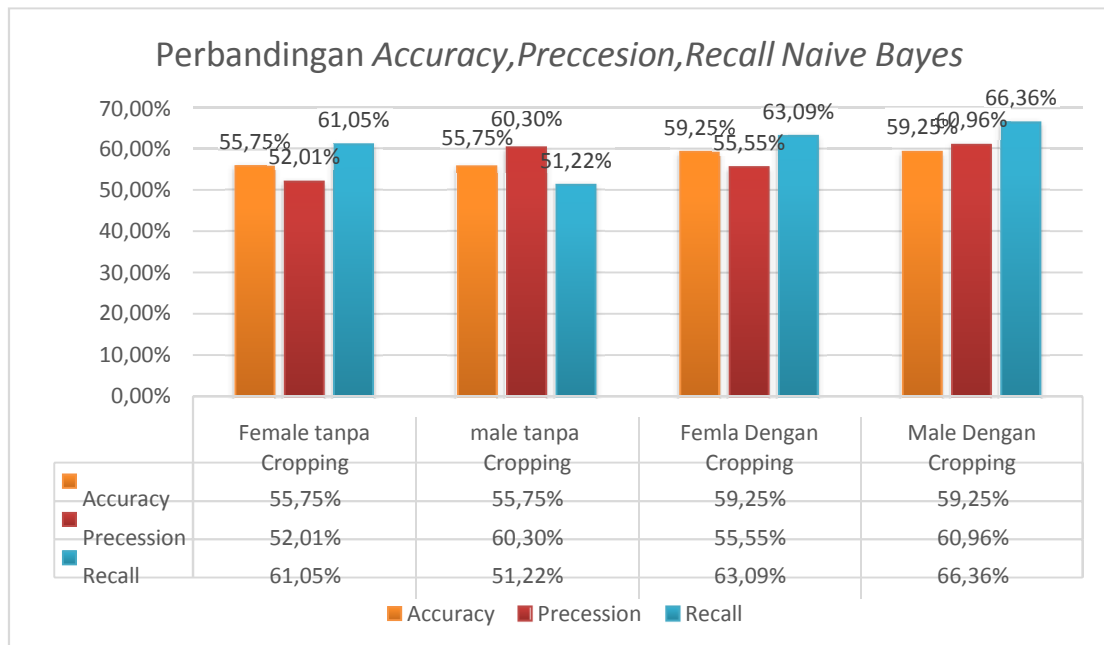


Gambar 6. Perbandingan Accuracy, Precision, Recall Dengan k=3

Dari hasil perbandingan yang telah ditunjukkan bahwa k=3 pengujian pada k-NN dengan menggunakan *Cropping* memiliki hasil Accuracy, lebih tinggi daripada tanpa *Cropping*, yaitu sebesar 65,70%, serta Precision dan Recall pada kelas Female dan male masing masing sebesar 62,99%, 68,06%, 63,23%, dan 67,83%

3.2.2 Analisis Skenario *Naïve Bayes*

Pada skenario *Naïve Bayes* yang dilakukan dengan menggunakan *Cropping* dan tanpa *Cropping* untuk klasifikasi mata. Perbandingan *accuracy*, *precision*, dan *recall* dapat dilihat pada Gambar 7.



Gambar 7. Perbandingan Accuracy, Precision, dan Recall pada Naive Bayes

Dari hasil perbandingan yang telah ditunjukkan bahwa pengujian metode Naive Bayes dengan menggunakan Cropping memiliki hasil Accuracy, lebih tinggi daripada tanpa Cropping, yaitu sebesar 59,25%, serta Precision dan Recall pada kelas Female dan male masing masing sebesar 52,01%, 63,09%, 60,96%, dan 57,77% .

4. KESIMPULAN

Berdasarkan hasil yang didapat dalam penelitian diatas, dapat disimpulkan sebagai berikut.

1. Pada metode k-NN dengan fitur HSV dengan Cropping mendapatkan hasil yang tinggi dalam menentukan gender berdasarkan citra mata didapat Accuracy dengan nilai 68,81%, Precision pada kelas Male dengan nilai 70,87% sedangkan kelas Female dengan nilai 66,42%, serta Recall pada kelas Male dengan nilai 70,99% sedangkan kelas Female dengan nilai 66,30% .
2. Pada metode Naive Bayes dengan fitur HSV dengan Cropping mendapatkan hasil yang lebih rendah dalam menentukan gender berdasarkan citra mata didapan Accuracy dengan nilai 62,58%, Precision pada kelas Male dengan nilai 65,65% sedangkan kelas Female dengan nilai 59,31% serta Recall pada kelas Male dengan nilai 61,67% sedangkan kelas Female dengan nilai 63,37%.
3. Dari jumlah dataset sebanyak 15,059 meliputi 7437 mata pria dan 7622 mata Wanita dapat disimpulkan bahwa metode k-NN lebih baik daripada metode Naive Bayes dengan tingkat akurasi sebesar 68,81% dibandingkan Naive Bayes yang lebih rendah dengan tingkat akurasi sebesar 62,58%.

DAFTAR PUSTAKA

- [1] Fakhri, D. M., *Analisis Gender & Transformasi Sosial.*, Yogyakarta: Pustaka Belajar, 2016.
- [2] Akhmad Rohim, "*Convolution Neural Network (CNN) Untuk Pengklasifikasian Citra Makanan Tradisional*," 2019.
- [3] D. Putra, *Pengolahan Citra Digital*, Yogyakarta: Andi, 2010.
- [4] Hidayatullah, P, *Pengolahan Citra Digital*, Bandung: Informatika, 2017.
- [5] Juditha, C., "*Gender dan Seksualitas Dalam Konstruksi Media Massa*," Vol. 4, p. 1(1), 2015.
- [6] Kusumadewi, S, "*Klasifikasi Status Gizi Menggunakan Naive Bayesian Classification*," Vol. 3(1), No. Commit, pp. 6-11, 2015.
- [7] Leidiyana, H, "*Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko*," 2017.
- [8] Myatt, Glenn J, "*Making Sense of Data, A Practical Guide to Exploratory Data Analysis and Data Mining*," in Hoboken, New Jersey, John Willey & Sons, 2007.
- [9] Prahudaya, T. Y, "*Metode Klasifikasi Mutu Jambu Biji Menggunakan KNN Berdasarkan Fitur Warna dan Tekstur*," *Jurnal Teknosains*, vol. 6(2), 2017.
- [10] Sebastian, G. I., "*Algoritme K-Nearest Neighbors Untuk Klasifikasi Jenis Makanan dari Citra Digital Dengan Local Binary Patterns dan Color Moments.*," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, Vol. 3(7), p. 6473–6479, 2019.
- [11] Wali Ja'far Shudiq, "*Penerapan K-Nearest Neighbor Berbasis Algoritma Genetika untuk Klasifikasi Mutu Padi Organik*," 2017.